CS 880: Pseudorandomness and Derandomization 2/20/2013

Lecture 9: Expanders

Instructors: Holger Dell and Dieter van Melkebeek Scribe: Adam Everspaugh

Expanders are directed or undirected multigraphs that are almost as well-connected as the complete graph but contain a lot fewer edges. We typically consider families of regular graphs of low (often constant) degree. The savings in edges are similar to the savings in randomness in PRGs, and in later lectures we will see how expanders can be used to construct various PRGs. In this lecture we discuss several measures of expansion and their relationships, and focus on the one known as spectral expansion. We also recap some linear algebra background for this and future lectures.

1 Measures of Expansion

Although we mostly deal with undirected multigraphs, the formal definitions in this section presume a directed multigraph $G = (V, E)$. In order to apply them to an undirected graph, we orient all edges both ways. We denote the neighborhood of a set of vertices $S \subseteq V$ as

$$\Gamma(S) \doteq \{v \in V | (\exists u \in S)(u, v) \in E\}.$$

The following quantities are commonly used to measure the well-connectedness of (multi)graphs. They are all closely related, as we will discuss further.

- Vertex Expansion
  This is the notion that gives expanders their name. Informally, we want any set S of vertices to expand a lot when we take its neighboring set $\Gamma(S)$. Formally, we require that $|\Gamma(S)| \geq c|S|$, where $c > 1$ is a parameter. Of course, this is only possible when the set S isn't too large to begin with, say $|S| \leq \alpha|V|$, where α is another parameter. This leads to the following definition: G has *vertex expansion* (α, c) if

  $$(\forall S \subset V \text{ with } |S| \leq \alpha|V|)\, |\Gamma(S)| \geq c|S|.$$

  In a complete digraph with self-loops every non-empty set S expands to all of V. If the maximum degree of G is d, the expansion c can be at most d. Typically, we want c to be some constant larger than 1, and α some positive constant.

  Note that vertex expansion does not require that $S \subseteq \Gamma(S)$, although that inclusion always holds when G has a self-loop at every vertex. Related to this, a slightly different way of formalizing that S expands is that $\Gamma(S)$ has a significant number of "new" vertices, i.e., $|\Gamma(S) \setminus S| \geq \gamma|S|$, where $\gamma > 0$ is a parameter. We say that G has *boundary expansion* (α, γ) if

  $$(\forall S \subset V \text{ with } |S| \leq \alpha|V|)\, |\Gamma(S) \setminus S| \geq \gamma|S|.$$

  Note that an (α, c) vertex expander is an (α, γ) boundary expander with $\gamma = c - 1$. The converse holds when G has a self-loop at every vertex.

- Edge Expansion
  Instead of looking at the new vertices in $\Gamma(S)$, we look at the number of edges leaving S, i.e., at $E(S, \overline{S}) \doteq E \cap S \times \overline{S}$. We say that that G has *edge expansion* (α, c) if

$$(\forall S \subset V \text{with} |S| \leq \alpha |V|)\, |E(S, \overline{S})| \geq c|S|.$$

  A closely related notion is that of *conductance*. If G is d-regular, the conductance of G equals the largest real Φ such that G has edge expansion $(\frac{1}{2}, \Phi \cdot d)$.

- Quasi-Randomness
  This is a notion from random graph theory. A random (di)graph on a given number of vertices is one where an independent fair coin is tossed to determine the presence of every individual edge. A graph G is called quasi-random if for any two (not necessarily disjoint) subsets of vertices, S and T, the fraction of edges that go from S to T is about the same as in a random digraph. More precisely, G is *ϵ-quasi-random* if

$$(\forall S, T \subset V) \left| \frac{|E(S,T)|}{|E|} - \mu(S)\mu(T) \right| \leq \epsilon,$$

  where $\mu(S) \doteq \frac{|S|}{|V|}$ denotes the relative size of S. Note that the complete digraph with self-loops is trivially ϵ-quasi-random for any $\epsilon \geq 0$.

  Phrased in terms of PRGs, G is ϵ-quasi-random if picking an edge from G is ϵ-pseudorandom for randomized algorithms that use two random vertices and check whether the first vertex is in S and the second one in T, where S and T can be any subsets. If G has small degree d, then this yields a nontrivial ϵ-PRG for this class of algorithms: Pick a vertex x from G uniformly at random, and then pick one of its neighbors y uniformly at random. In order to generate $2 \log |V|$ bits, this PRG only needs $\log |V| + \log d$ truly random bits. We will use this PRG as a building block later on.

- Rapid Mixing
  This is a notion from the theory of Markov processes. We consider the Markov process of a random walk on G: Start at some vertex, and repeatedly move to a neighbor chosen uniformly at random. We require the process to mix rapidly, i.e., to quickly converge to a unique invariant distribution for every possible start vertex. For the complete digraph with self-loops, the process converges in a single step to the uniform distribution. In the next section we provide some background on Markov processes and the notion of rapid mixing.

- Spectral Expansion
  This notion forms the basis for our definition of expanders. It is linear-algebraic in nature and also involves the above random walk. We present it after our discussion of rapid mixing and a recap of the relevant linear algebra.

2 Rapid Mixing

A *Markov process* is a memoryless randomized process. Specifically, it is a randomized process where the probability distribution for the next state is determined only by the current state – with

no memory of the prior states. If the state space is finite, say of size N, the process can be specified by the $N \times N$ transition matrix M with $M_{ij} = \Pr[\text{next state is } i \,|\, \text{current state is } j]$.

To run the process we also need an initial state distribution $\pi^{(0)}$, which we view as a column vector of length N with $\pi^{(0)}(i) = \Pr[\text{initial state is } i]$. The column vector $\pi^{(t)}$ describing the state distribution after t steps of the process is then given by the matrix-vector product $\pi^{(t)} = M^t \pi^{(0)}$.

A *random walk* on a directed multigraph is such a process. If there are no parallel edges, we can write

$$M_{ij} = \begin{cases} \frac{1}{\text{outdeg}(j)} & \text{if} (j,i) \in E \\ 0 & \text{otherwise.} \end{cases}$$

If there are parallel edges, then the numerator 1 is replaced by the number of edges from j to i. If every vertex has outdegree d, we can write M as $M = \frac{1}{d} A^T$, where A denotes the adjacency matrix of G (with multiplicities).

An *invariant distribution* (also called a stationary distribution) of a Markov process is an initial distribution π that remains unchanged throughout the process, i.e., such that $M\pi = \pi$. A Markov process always has an invariant distribution, but it may not be unique. The latter depends on the structure of the tree of strongly connected components (SCCs) of the underlying multigraph of transitions with nonzero probabilities, which can be seen as follows. Note that for a random walk on G the underlying multigraph is G itself.

- Every vertex in the support of an invariant distribution has to belong to an SCC that is a leaf in the tree of SCCs, i.e., an SCC from which no other SCC can be reached. This is because for all other SCCs one step of the process permanently leaks a nonzero amount of its probability mass to SCCs further down the tree.

- For every leaf SCC C there is an invariant distribution whose support coincides with C. To see this, first note that the restriction of M to rows and columns in C defines a Markov process M' on C. Since $\vec{1}M' = \vec{1}$, M' has a left eigenvector with eigenvalue 1, and thus also a (right) eigenvector π with eigenvalue 1. This vector is not identically zero and satisfies $M\pi = \pi$. It remains to show that, up to a scalar, π is a probability distribution with support C, i.e., that all components of π have the same sign. Since C is strongly connected, it suffices to argue the following claim for $i, j \in C$: If $M_{ij} > 0$ then the sign of π_j equals the sign of π_i or $\pi_j = 0$. If $\pi_i = 0$ for some $i \in C$, then the claim implies that $\pi_j = 0$ for all $j \in C$ from which i can be reached, which is all of C. Since π is not identically zero, we conclude that $\pi_i \neq 0$ for every $i \in C$. By the same token, the claim then shows that the sign of π_j is the same for every $j \in C$.

  We argue the claim by contradiction. Suppose that $M_{ij} > 0$, $\pi_j \neq 0$, and the sign of π_i differs from the sign of π_j. Since $\pi = M\pi$, we have that

  $$\pi_i = M_{ij}\pi_j + \sum_{k \neq j} M_{ik}\pi_k.$$

  Under the given sign conditions, this equality can only hold if at least one of the terms on the right-hand side with $k \neq j$ has the opposite sign of the first term, which implies that

  $$|\pi_i| < \sum_k M_{ik}|\pi_k|.$$

If we define the probability distribution π' on C by $\pi'_i = |\pi_i| / \sum_{j \in C} |\pi_j|$ for $i \in C$, this further implies that $1 = \sum_i \pi'_i < \sum_i (M\pi')_i$. Thus $M\pi'$ is not a probability distribution, which contradicts the fact that M transforms every probability distribution into a probability distribution.

The two points combined show that a Markov process has a unique invariant distribution iff the tree of SCCs has a single leaf. For a random walk on an undirected multigraph G this is the case iff G is connected.

We call a randomized process *mixing* if it converges to a unique distribution that is independent of the initial distribution. For a Markov process to be mixing, it has to have a unique invariant distribution but that is not enough. For example, consider the case where a vertex v in the unique leaf SCC has a nontrivial period. By this we mean that there exists an integer $p > 1$ such that the Markov process starting from state v only has a positive probability of being in state v at time steps that are multiples of p. In that case the Markov process does not converge. We leave it as an exercise to show that this is the only bad case.

Exercise 1. *Show that a Markov process is mixing iff (i) it has a unique invariant distribution, and (ii) no state in the unique leaf SCC has a nontrivial period.*

For a random walk on an undirected multigraph G, the only possible nontrivial period p is 2, as it is always possible to return from a vertex v itself in two steps. It can be seen that a vertex in a leaf SCC C has period 2 iff C is bipartite. Thus, the random walk process on an undirected multigraph G is mixing iff G is connected and not bipartite.

The mixing is called *rapid* if the convergence to the unique invariant distribution π happens fast. More precisely, for some slowly growing function t of $|V|$ and $\frac{1}{\epsilon}$, we require that $d_{\text{stat}}(\pi^{(t)}, \pi) \leq \epsilon$ for every initial distribution $\pi^{(0)}$ and every positive ϵ. Typically, we aim for t logarithmic in $|V|$ and $\frac{1}{\epsilon}$.

For the purposes of constructing PRGs, we are mostly interested in the case where the unique invariant distribution is *uniform*. For an undirected multigraph G the distribution that assigns a vertex a probability proportional to its degree is always invariant. This distribution is uniform iff G is regular. For regular directed multigraphs (where all vertices have the same outdegree as well as indegree), the uniform distribution is also always invariant. This is the reason why in the sequel we focus on *regular* (directed) multigraphs.

3 Linear Algebra Background

Before moving on to the notion of spectral expansion, we first review some facts from linear algebra that will be useful for the rest of this lecture as well as for later lectures.

3.1 Vector Norms

We work with N-dimensional vectors over $\mathbb{R}$ or $\mathbb{C}$. For $x, y \in \mathbb{C}^N$, we define their inner product as $\langle x, y \rangle \doteq \sum_{i=1}^{N} x_i \overline{y_i}$, where $\overline{y_i}$ denotes the complex conjugate of y_i. In the case of real vectors the complex conjugation can be dropped. Sometimes we also use the scaled inner product $(x, y) \doteq \frac{1}{N} \langle x, y \rangle$. These definitions agree with the ones we saw in Lecture 7 for the inner product of functions $f : G \to \mathbb{C}$ by viewing a vector $x \in \mathbb{C}^N$ as a function $f : [N] \to \mathbb{C}$ with $f(i) = x_i$.

A *norm* on a vector space is a function $\|\cdot\|$ from the vector space to $\mathbb{C}$ satisfying the following properties for all vectors x, y and scalar $\alpha \in \mathbb{C}$:

- Non-negativity: $\|x\|$ is a nonnegative real, and $\|x\| = 0$ if and only if x is the zero vector.
- Linearity: $\|\alpha x\| = |\alpha| \cdot \|x\|$.
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$.

Exercise 1 from Lecture 7 shows that $\|x\|_2 \doteq \sqrt{\langle x, x\rangle}$ defines a valid norm on $\mathbb{C}^N$. More generally, so does $\|x\|_p \doteq \sqrt[p]{\sum_{i=1}^N |x_i|^p}$ for every $p \in [1, \infty]$, where $\|x\|_\infty \doteq \lim_{p\to\infty} \|x\|_p = \max_i |x_i|$. $\|\cdot\|$ is referred to as the p-norm. In addition to the case $p = 2$, we will only use $p = 1$ and $p = \infty$, for which the above properties are straightforward to verify.

If x represents a probability distribution π, then the 1-norm is always 1, and the ∞-norm equals the maximum probability of any of the outcomes. The square of the 2-norm represents the *collision probability* of the distribution, i.e., the probability that two independent draws yield the same outcome.

In the context of probability the 1-norm is most interesting because the statistical distance between two distributions equals half the 1-norm of their difference. However, in a linear-algebraic context the 2-norm is often easier to handle because of the underlying inner product. Thus, we often switch between the 1-norm and the 2-norm, and rely on the following inequalities.

We have that $\|x\|_1 \geq \|x\|_2 \geq \|x\|_\infty$, and the equalities hold iff x has at most one nonzero component. A notable inequality in the other direction follows from the *Cauchy-Schwarz inequality*, which states that

$$|\langle x, y\rangle| \leq \|x\|_2 \cdot \|y\|_2,$$

where the equality holds iff x and y are identical up to a multiplicative scalar. By setting $y_i = \overline{x_i}/|x_i|$ if $x_i \neq 0$ and $y_i = 0$ otherwise, the Cauchy-Schwarz inequality tells us that

$$\|x\|_1 \leq \sqrt{\#\{i \in [N] \,:\, x_i \neq 0\}} \cdot \|x\|_2, \tag{1}$$

where the equality holds iff all nonzero components of x have the same absolute value. In particular, $\|x\|_1 \leq \sqrt{N}\|x\|_2$ always holds, and equality holds iff all components of x have the same absolute value.

3.2 Matrix Norms

Given a vector norm $\|\cdot\|$, we can define a matrix norm as the maximum stretch with respect to $\|\cdot\|$ of a nonzero-vector under the linear operation defined by the matrix: For $M \in \mathbb{C}^{N\times N}$

$$\|M\| \doteq \max_{x\in\mathbb{C}^N\setminus\{0^N\}} \frac{\|Mx\|}{\|x\|}. \tag{2}$$

Exercise 2. *Show that:*

- (2) *defines a valid matrix norm,*
- $\|M\| = \max_{\|x\|=1} \|Mx\|$,

- $\|Mx\| \leq \|M\| \cdot \|x\|$, *and*
- $\|MM'\| \leq \|M\| \cdot \|M'\|$,

where x *ranges over* $\mathbb{C}^N$, *and* M *and* M' *over* $\mathbb{C}^{N \times N}$.

Of particular interest to us are the matrix norms induced by the p-norm for $p \in \{1, 2, \infty\}$.

Exercise 3. *Show that:*

- $\|\cdot\|_1$ *induces the column-sum norm, i.e.,* $\|M\|_1 = \max_{j \in [N]} \sum_{i=1}^{N} |M_{ij}|$, *and*
- $\|\cdot\|_\infty$ *induces the row-sum norm, i.e.,* $\|M\|_\infty = \max_{i \in [N]} \sum_{j=1}^{N} |M_{ij}|$,

where M *ranges over* $\mathbb{C}^{N \times N}$.

In particular, if M represents the transition matrix of a Markov process, then $\|M\|_1 = 1$.

3.3 Spectral Decompositions

A spectral decomposition of a vector involves the *spectrum* of a matrix M, a term that either refers to the (multi)set of eigenvalues of M or the (multi)set of singular values of M.

For a *symmetric* real matrix $M \in \mathbb{R}^{N \times N}$, all eigenvalues are real, eigenvectors belonging to distinct eigenvalues are orthogonal, and M has a set of real eigenvectors that forms an orthogonal basis for $\mathbb{R}^N$: There exist $\lambda_i \in \mathbb{R}$ and $v_i \in \mathbb{R}^N \setminus \{0^N\}$ for $i \in [N]$ such that

- $Mv_i = \lambda_i v_i$ for every $i \in [N]$, and
- $\langle v_i, v_j \rangle = 0$ for every $i, j \in [N]$ with $i \neq j$.

This implies that every $x \in \mathbb{R}^N$ can be decomposed in a unique way as

$$x = \sum_{i=1}^{N} \xi_i v_i, \tag{3}$$

where $\xi_i \in \mathbb{R}$, namely $\xi_i = \langle x, v_i \rangle / \langle v_i, v_i \rangle$. Such an expression is called an *eigenvalue decomposition* of x with respect to M.

Matrices that have a full orthogonal basis of eigenvectors are called *normal*. Although there are non-symmetric real matrices that are normal, not every real matrix has a full basis of eigenvectors, let alone an orthogonal one. However, every real matrix has a full orthogonal basis of singular vectors, obtained as follows.

For a given real matrix $M \in \mathbb{R}^{N \times N}$, consider $M' \doteq M^T M$. Note that M' is a real symmetric matrix, and therefore has a full orthogonal basis of real eigenvectors v_i, $i \in [N]$, with corresponding eigenvalues λ_i. The v_i's are referred to as the *singular vectors* of M. Note that the eigenvalues λ_i of M' are nonnegative reals, as

$$\lambda_i = \frac{v_i^T M' v_i}{v_i^T v_i} = \frac{(Mv_i)^T (Mv_i)}{v_i^T v_i} = \frac{\|Mv_i\|_2^2}{\|v_i\|_2^2} \in [0, \infty].$$

The values $\sigma_i \doteq \sqrt{\lambda_i}$ are therefore well-defined, and are called the *singular values* of M. An eigenvalue decomposition (3) of x with respect to M' is called a *singular value decomposition* of x with respect to M. It follows that a singular value decomposition of x exists with respect to any real matrix M. Moreover, in the case where M is a real *symmetric* matrix, $M' = M^2$, the singular vectors of M coincide with the eigenvectors of M, the singular values of M are the absolute values of the corresponding eigenvalues of M, and the singular value decomposition coincides with the eigenvalue decomposition.

4 Spectral Expansion

We are now ready to introduce the notion of spectral expansion, which forms the basis for our definition of an expander. Although spectral expansion can be defined more generally, we develop it only for *regular* directed multigraphs. This is because we only need the regular case, and the development for that case is simpler. In fact, the case of regular *undirected* multigraphs is even simpler, so we present that first.

4.1 Regular Undirected Case

Let G be a d-regular undirected multigraph on N vertices, and let M_G represent the transition matrix of the random walk on G. Note that M_G is a symmetric real matrix, and therefore all of its eigenvalues are real. We already argued that 1 is an eigenvalue of M_G, and the uniform distribution U_N is a corresponding eigenvector. Moreover, all eigenvalues are at most 1 in absolute value. This follows from the fact that as the transition matrix of a Markov process, $\|M_G\|_1 \leq 1$: For any eigenvector v with eigenvalue λ, we have that

$$|\lambda|\|v\|_1 = \|\lambda v\|_1 = \|M_G v\|_1 \leq \|M\|_1 \cdot \|v\|_1 = \|v\|_1,$$

or more explicitly:

$$|\lambda| \sum_{i=1}^{N} |v_i| = \sum_{i=1}^{N} |(M_G v)_i| = \sum_{i=1}^{N} |\sum_{j=1}^{N} (M_G)_{ij} v_j| \leq \sum_{i=1}^{N} \sum_{j=1}^{N} (M_G)_{ij} |v_j| = \sum_{j=1}^{N} (\sum_{i=1}^{N} (M_G)_{ij}) |v_j| = \sum_{j=1}^{n} |v_j|,$$

which implies that $|\lambda| \leq 1$. Thus, we can order the N eigenvalues of M_G (including multiplicities) in order of non-increasing absolute value starting with $\lambda_1 = 1$:

$$1 = \lambda_1 \geq |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_N| \geq 0.$$

The gap between the top two absolute values defines the spectral expansion.

Definition 1 (Spectral Expansion – Undirected). *An undirected regular multigraph G has* spectral expansion γ *if*

$$\lambda(G) \doteq |\lambda_2| \leq 1 - \gamma,$$

where λ_2 denotes the eigenvalue of M_G that is the second largest in absolute value. We call $1 - |\lambda_2|$ the spectral gap of G.

The eigenvalue decomposition yields the following alternate characterization.

Proposition 1. *For an undirected regular multigraph G,*

$$\lambda(G) = \max_{0^N \neq x \perp U_N} \frac{\|M_G x\|_2}{\|x\|_2}.$$

Proof. Consider the eigenvalue decomposition (3) with respect to M_G of a nonzero vector $x \in \mathbb{R}^N$ that is orthogonal to U_N. Since $v_1 = U_N$, we have that $\xi_1 = 0$, and by the orthogonality of the v_i's

$$\|M_G x\|_2^2 = \|\sum_{i=2}^{N} \lambda_i \xi_i v_i\|_2^2 = \sum_{i=1}^{N} |\lambda_i|^2 |\xi_i|^2 \|v_i\|_2^2 \leq \max_{i=1}^{N}(|\lambda_i|^2) \sum_{i=1}^{N} |\xi_i|^2 \|v_i\|_2^2 = |\lambda_2|^2 \|x\|_2^2 = \lambda(G)^2 \|x\|_2^2.$$

Moreover, equality holds for $x = v_2$, which is orthogonal to U_N. The proposition follows by taking square roots on both sides and dividing by $\|x\|_2$. □

The value of $\lambda(G)$ ranges over $[0, 1]$. It equals 0 iff M_G has rank 1. This is the case for the complete digraph with self-loops, and there is no other regular graph for which this is the case. At the other end, $\lambda(G) = 1$ iff G is disconnected or has a bipartite component. This follows from the next properties, which are left as an exercise.

Exercise 4. *Show that:*

- *the multiplicity of 1 as an eigenvalue of M_G equals the number of connected components of G, and*
- *-1 is an eigenvalue of M_G iff G has a bipartite component.*

4.2 Regular Directed Case

Recall that a directed multigraph G is d-regular if every vertex has indegree and outdegree d. We consider the singular values of M_G, which are the roots of the eigenvalues of $M_G^T M_G$. Note that the regularity of G implies that M_G^T is the transition matrix of a random walk, namely on the digraph obtained by reversing all the edges of G. This implies that $M_G^T M_G$ is the transition matrix of a Markov process, and therefore the eigenvalues of $M_G^T M_G$ are at most 1 in absolute value, so the singular values of M_G are at most 1. Moreover, U_N is an invariant distribution of the random walks on G and the reverse of G, so 1 is a singular value of M_G. Thus, we can order the N singular values of M_G (including multiplicities) in non-increasing order starting with $\sigma_1 = 1$:

$$1 = \sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_N \geq 0.$$

The gap between the top two singular values defines the spectral expansion.

Definition 2 (Spectral Expansion – Directed). *An directed regular multigraph G has* spectral expansion γ *if*

$$\lambda(G) \doteq \sigma_2 \leq 1 - \gamma,$$

where σ_2 denotes the second largest singular value of M_G. We call $1 - \sigma_2$ the spectral gap of G.

Note that for an undirected regular multigraph G, directing the edges both ways yields a directed regular multigraph G' with $M_G = M_{G'}$, and Definitions 1 and 2 agree. This follows from the connections between the singular values and eigenvalues of symmetric matrices. They imply that $\sigma_i(G') = |\lambda_i(G)|$ for every $i \in [N]$, and thus that $\lambda(G) = \lambda(G')$.

The alternate characterization given by Proposition 1 generalizes.

Proposition 2. *For a directed regular multigraph G,*

$$\lambda(G) = \max_{0^N \neq x \perp U_N} \frac{\|M_G x\|_2}{\|x\|_2}.$$

The proof uses the singular value decomposition instead of the eigenvalue decomposition.

Proposition 2 forms the basis for the following interpretation of the spectral gap as a measure for the similarity to the complete digraph with self-loops.

Proposition 3. *Let G be a regular directed multigraph on N vertices, and K the complete directed graphs with self-loops on N vertices. G has spectral gap at least γ iff there exists a matrix $\Delta \in \mathbb{R}^{N \times N}$ with $\|\Delta\|_2 \leq 1$ such that*

$$M_G = \gamma \cdot M_K + (1 - \gamma) \cdot \Delta.$$

The proposition states that one step of a random walk on G can be viewed as taking one step of a random walk on K with probability γ, and doing "something else" with probability $1 - \gamma$, where "something else" does not need to be a Markov process but cannot affect the state of the system by too much because $\|\Delta\|_2 \leq 1$.

Proof. Consider $M \doteq M_G - \gamma M_K$. Since all of M_G, M_K, and their transposes leave U_N invariant, U_N is a singular vector of M (with singular value 1). It follows that both of the following spaces are closed under the operation of M:

- $U_N^{\parallel} \doteq \{x \in \mathbb{R}^N : x \parallel U_N\}$
  For $x^{\parallel} \parallel U_N$, we have that $Mx^{\parallel} = (1 - \gamma)x^{\parallel}$.
- $U_N^{\perp} \doteq \{x \in \mathbb{R}^N : x \perp U_N\}$
  For $x^{\perp} \perp U_N$, we have that $M_K x^{\perp} = 0$ and therefore $Mx^{\perp} = M_G x^{\perp}$.

As both spaces are orthogonal complements, it follows that each $x \in \mathbb{R}^n$ can be decomposed as $x = x^{\parallel} + x^{\perp}$ with $x^{\parallel} \parallel U_N$ and $x^{\perp} \perp U_N$ such that $\|x^{\parallel}\|_2^2 + \|x^{\perp}\|_2^2 = \|x\|_2^2$ and

$$\|Mx\|_2^2 = \|Mx^{\parallel}\|_2^2 + \|Mx^{\perp}\|_2^2 = (1 - \gamma)^2 \|x^{\parallel}\|_2^2 + \|M_G x^{\perp}\|_2^2.$$

It follows that $\|M\|_2 = \max((1 - \gamma), \max_{0^n \neq y \perp U_N}(\|M_G y^{\perp}\|_2 / \|y\|_2))$, and by Proposition 2 that $\|M\|_2 = \max(1 - \gamma, \lambda(G))$. We conclude that $\|M\|_2 \leq 1 - \gamma$ iff $\lambda(G) \leq 1 - \gamma$, which is equivalent to the statement of the proposition. □

5 Properties of Expanders

Spectral expansion and all of the other measures of expansion mentioned before are equivalent up to some loss in parameters. We now analyze the direction that matters to us, namely that our definition of expanders based on spectral expansion implies all other expansion properties.

5.1 Rapid Mixing

The connection with rapid mixing is the most immediate one, and is tight.

Let G be a regular directed multigraph, and M_G the transition matrix for the random walk on G. Because of regularity, the uniform distribution U_N is invariant under M_G.

Let $\pi^{(0)}$ denote the distribution at the start of the random walk. Measuring distance using the 2-norm we have:

$$\|\pi^{(t)} - U_N\|_2 = \|M_G^t \pi^{(0)} - U_N\|_2 = \|M_G^t(\pi^{(0)} - U_N)\|_2 \leq \lambda(G)^t \|\pi^{(0)} - U_n\|_2,$$

where the inequality follows from t repeated applications of Proposition 2 and the fact that for any distribution D, $(D - U_N) \perp U_N$. Moreover, the inequality becomes an equality for $\pi^{(0)}$ of the form $\pi^{(0)} = U_N + \alpha \cdot v_2$, which is a probability distribution for some sufficiently small positive real α.

In terms of statistical distance, this implies that

$$d_{\text{stat}}(\pi^{(t)}, U_N) = \frac{1}{2}\|\pi^{(t)} - U_N\|_1 \leq \frac{1}{2}\sqrt{N}\|\pi^{(t)} - U_N\|_2 \leq \frac{1}{2}\sqrt{N}\lambda(G)^t\|\pi^{(0)} - U_N\|_2 \leq \frac{1}{2}\sqrt{N}\lambda(G)^t,$$

where the last step follows because for any distribution D

$$\|D - U_N\|_2^2 = \|D\|_2^2 - \|U_N\|_2^2 \leq \|D\|_1^2 = 1.$$

Thus, $d_{\text{stat}}(\pi^{(t)}, U_N) \leq \epsilon$ for $t \geq \frac{\log(\sqrt{N}/\epsilon)}{\log(1/\lambda(G))}$, which is logarithmic in N/ϵ when $\lambda(G)$ is bounded from above by a constant less than 1, i.e., when the spectral gap of G is at least some positive constant.

5.2 Quasi-Randomness

A regular directed multigraph G is ϵ-quasi-random for $\epsilon = \lambda(G)$. This follows from the next result, which is often referred to as the expander mixing lemma.

Theorem 4 (Expander Mixing Lemma). *Let $G = (V, E)$ be a d-regular directed multigraph. For all $S, T \subseteq V$*

$$\left|\frac{|E(S,T)|}{|E|} - \mu(S)\mu(T)\right| \leq \lambda(G) \cdot \sqrt{\mu(S)(1-\mu(S))}\sqrt{\mu(T)(1-\mu(T))},$$

where $\mu(S) \doteq \frac{|S|}{|V|}$ denotes the relative size of S.

Proof. We associate V with $[N]$ and let I_S and I_T denote the characteristic vectors of S and T, respectively, i.e., I_S is the column vector of dimension N with the i-th component equal to 1 if $i \in S$, and 0 otherwise.

We first express $|E(S,T)|$ in linear-algebraic terms using M_G:

$$|E(S,T)| = d \cdot I_T^T M_G I_S. \tag{4}$$

To see this, note that $dM_G = A^T$, where A denotes the adjacency matrix of G (with multiplicities). Therefore,

$$d \cdot I_T^T M_G I_S = I_T^T A^T I_S = \sum_{i,j=1}^{N} (I_T)_j A_{ij} (I_S)_i = \sum_{i \in S}\sum_{j \in T} A_{ij} = |E(S,T)|.$$

In order to exploit the spectral gap of G in evaluating (4), we decompose I_S and I_T into their components parallel to U_N and orthogonal to U_N. For I_S note that $\langle I_S, U_N\rangle = \mu(S)$, $\langle U_N, U_N\rangle = 1/N$, and $\langle I_S, I_S\rangle = |S|$. It follows that

$$I_S = |S|U_N + I_S^\perp,$$

where

$$\|I_S^\perp\|_2^2 = \|I_S\|_2^2 - \||S|U_N\|_2^2 = |S| - |S|^2\frac{1}{N} = |S|\cdot(1-\mu(S)). \tag{5}$$

We decompose I_T in a similar way. Using those decompositions we obtain

$$\begin{aligned} I_T^T M_G I_S &= (|T|U_N + I_T^\perp)^T M_G(|S|U_N + I_S^\perp) \\ &= (|T|U_N + I_T^\perp)^T(|S|U_N + M_G I_S^\perp) \\ &= (|T|U_N)^T(|S|U_N) + (I_T^\perp)^T M_G I_S^\perp, \end{aligned}$$

where the cross terms disappeared by orthogonality and the fact that M_G maps $I_S^\perp$ to a vector orthogonal to U_N. As the first term equals $N\mu(S)\mu(T)$, using (4) and the fact that $|E| = dN$, we obtain

$$\frac{|E(S,T)|}{|E|} = \mu(S)\mu(T) + \frac{1}{N}(I_T^\perp)^T M_G I_S^\perp.$$

We conclude that

$$\begin{aligned} \left|\frac{|E(S,T)|}{|E|} - \mu(S)\mu(T)\right| &= |\frac{1}{N}(I_T^\perp)^T M_G I_S^\perp| = \frac{1}{N}|\langle I_T^\perp, M_G I_S^\perp\rangle| \\ &\le \frac{1}{N}\|I_T^\perp\|_2\cdot\|M_G I_S^\perp\|_2 \\ &\le \frac{1}{N}\|I_T^\perp\|_2\cdot\lambda(G)\cdot\|I_S^\perp\|_2 = \lambda(G)\cdot\sqrt{\mu(S)(1-\mu(S))}\sqrt{\mu(T)(1-\mu(T))}, \end{aligned}$$

where we used Proposition 2 and applied (5) to S and T. □

Instead of decomposing the vectors I_S and I_T as in the proof of Theorem 4, we can alternately decompose the matrix M_G as in Proposition 3. The derivation is somewhat shorter but also yields a somewhat weaker result.

We follow the outline and notation from the proof of Theorem 4. Using (4) and the decomposition $M_G = \gamma M_K + (1-\gamma)\Delta$ from Proposition 3 with $\gamma = 1-\lambda(G)$ we have that

$$|E(S,T)| = dI_T^T M_G I_S = (1-\lambda(G))dI_T^T M_K I_S + \lambda(G)dI_T^T\Delta I_S^T.$$

Applying (4) with $G = K$ tells us that $NI_T^T M_K I_S = |S||T|$, which leads to

$$\frac{|E(S,T)|}{|E|} = (1-\lambda(G))\mu(S)\mu(T) + \lambda(G)\frac{1}{N}I_T^T\Delta I_S^T,$$

whence

$$\left|\frac{|E(S,T)|}{|E|} - \mu(S)\mu(T)\right| \le \lambda(G)\left(\mu(s)\mu(T) + \frac{1}{N}|I_T^T\Delta I_S|\right).$$

Since $\|\Delta\|_2 \le 1$, we have that

$$|I_T^T\Delta I_S| = |\langle I_T, \Delta I_S\rangle| \le \|I_T\|_2\cdot\|\Delta I_S\|_2 \le \|I_T\|_2\cdot\|\Delta\|_2\cdot\|I_S\|_2 \le \|I_T\|_2\cdot\|I_S\|_2 = \sqrt{|S||T|},$$

which leads to the conclusion

$$\left|\frac{|E(S,T)|}{|E|} - \mu(S)\mu(T)\right| \leq \lambda(G)\sqrt{\mu(S)\mu(T)}\left(1+\sqrt{\mu(S)\mu(T)}\right).$$

Note that this upper bound is somewhat weaker than the one given in Theorem 4.

5.3 Edge Expansion

Edge expansion follows immediately from quasi-randomness. Picking $T = \overline{S}$ in Theorem 4 yields that

$$\left|\frac{|E(S,\overline{S})|}{dN} - \mu(S)\mu(\overline{S})\right| | \leq \lambda(G)\mu(S)\mu(\overline{S}),$$

whence

$$|E(S,\overline{S})| \geq (1-\lambda(G))d|S|\mu(\overline{S}).$$

This shows that G has edge expansion $(\alpha, (1-\lambda(G))(1-\alpha)d)$ for every $\alpha \in [0,1]$, and conductance $\Phi \geq \frac{1-\lambda(G)}{2}$.

5.4 Vertex Expansion

For a d-regular directed multigraph G, the result for edge expansion implies that G has boundary expansion $(\alpha, (1-\lambda(G))(1-\alpha))$ for every $\alpha \in [0,1]$. This is because every "new" vertex can induce at most d edges in $E(S,\overline{S})$.

If G has a self-loop at every vertex, it follows that G has vertex expansion $(\alpha, 1+(1-\lambda(G))(1-\alpha))$. Vertex expansion follows from spectral expansion in general; we refer to [Tan84] for an argument that works in the general case.

In particular, a family of expanders G with constant degree d and $\lambda(G) \leq \lambda$ for some constant $\lambda < 1$ has vertex expansion $(\frac{1}{2}, c)$ for some constant $c > 1$. It follows that the *diameter* of G is $O(\log N)$. This is because when we iteratively take the neighbor set starting from the singleton $\{u\}$ for any fixed vertex u, the neighbor set keeps growing by a factor of c or more until it occupies more than half the vertices. The latter happens after no more than $s \doteq \log_c(N)$ iterations. Thus, for any two such vertices, u and v, these sets have to overlap after s steps, so there is a path of length at most $2s$ between u and v. As u and v are arbitrary, this means that the diameter of G is no more than $2s = O(\log N)$.

References

[Tan84] Michael Tanner. Explicit concentrators from generalized N-gons. *SIAM Journal on Algebraic Discrete Methods*, 5:287–293, 1984.